Supporting Public Health Surveillance through the National Electronic Disease Surveillance System (NEDSS)

Methods for conducting public health surveillance may often differ considerably by program and disease. Regardless of these differences, however, all surveillance activities share many common practices in the way data are **collected**, **managed**, **transmitted**, **analyzed, accessed** and **disseminated**. The **National Electronic Disease Surveillance System (NEDSS)** will, primarily through the creation of **standards**, facilitate the handling of data through each of these steps. As described below, different interrelated activities supporting NEDSS will offer significant improvements in the way public health surveillance is conducted at the Federal, State, and local level. The long-term vision for NEDSS is that of complementary electronic information systems that automatically gather health data from a variety of sources on a real-time basis; facilitate the monitoring of the health of communities; assist in the ongoing analysis of trends and detection of emerging public health problems; and provide information for setting public health policy.

Data Collection

Cases of a disease or other condition of interest are primarily identified in the medical care system. Once identified, these cases are typically reported to a local health department, often using paper-based data collection forms. At the local health department, forms may be entered into a computerized electronic data management system and transmitted to the State, or they may be copied, filed at the local level and then sent directly to the State where they are manually entered into the State health department's electronic system. These reporting processes are generally the same, regardless of the disease or condition that is being reported.

There are a number of problems that can arise during the reporting process. These problems, in turn, often place a large burden on medical care staff who have responsibility for disease reporting. For example, cases may be reported from a variety of sources from within the health care setting, such as clinical laboratories and private physicians. Physicians and laboratory supervisors and their office staff are already overworked. Nevertheless, it is often left up to health care providers to determine if a case meets public health surveillance case definitions and to figure out how to fill out the wide variety of forms produced by CDC and health departments. They may also spend significant time tracking down patient records in response to requests for more information from the health department.

NEDSS will facilitate the collection of case report forms from providers in two important ways. First, **standards** are being developed to assure uniform data collection practices across the nation. The **public health data model** and the **CIPHER** (Common Information for Public Health Electronic Reporting) guidelines will recommend, for example, a minimum set of demographic data that should be collected as part of the routine surveillance. In addition, the **CIPHER** guidelines will provide a consistent method for coding data on the data collection forms. It is expected that standardizing data collection forms should ease the burden on physicians and their staff by providing forms that are similar in appearance and that do not require that someone figure out for each specific form where information is located and how it should be coded.

Second, NEDSS will include recommended **standards** that can be used for the electronic reporting of surveillance data. Specifically, NEDSS will recommend a **standard data architecture** and **electronic data interchange format** to allow computer systems to generate automatically electronic case reports that can be sent to local or State health departments. These types of standards would both ease the burden on large organizations that already have computerized data systems (such as regional laboratories, hospitals, managed care organizations) and would ensure that all cases that are in the provider's data systems are being reported to public health.

Data management issues: a) multiple case reports

To whom cases should be and are reported is often unclear. For example, a physician reporting a case would likely send the form to the county health department. A State or regional laboratory may also report the same case to the State health department. Given the number of potential sources of information regarding a single patient, the possibility exists that persons may be entered into the system more than once or may have discrepant data reported about them on the multiple reports. In addition, undoing these duplicate records after the reports have been received at the health department (often weeks or months later) is more cumbersome than detecting those duplicate records and consolidating them prior to entry into the system data base. For example, the original paper records needed to resolve discrepant data may not be easily retrievable or may be lost.

To address this problem, NEDSS will identify **standard software components/tools** that may be used at the local and State health department levels to detect duplicate reports based on a person's demographic data (e.g., name, address, date of birth, sex). This process is known as **registry matching** (also referred to as **"record matching"**).   As paper forms are entered into the electronic system, the database of prior records would be scanned and potential duplicates identified. Next, data entry operators could decide whether to enter the particular case as a new report or to update the record already present.

The need for an automated system of **registry matching** is even more important as we move toward increasing reliance on automated electronic case reporting. While paper forms are generally handled one at time for entry and processing, electronic records are usually received in bulk and are processed together. The **record matching software** must be able to reliably determine which records are new and which should update existing reports. In addition, the software must also be capable of detecting instances of discrepant data, and, as deemed appropriate by the programs, it must be able to resolve those discrepancies. Finally, the tool must provide a mechanism for saving enough information on the individual reports in electronic archives if necessary, so that if at a later date two records were found to be merged inappropriately, the original records can be restored.

Data Management Issues: b) Data Entry at the Health Departments

The multiple data entry systems that CDC has created for local and State health department use have led to many complaints. Chief among these is that that data for an individual person must

be entered into multiple, disparate systems. Given recent advances in technology, this is an unnecessary and burdensome step. A second common complaint is that each of the CDC-provided systems works differently, so that using more than one of them is onerous and time-consuming. An analogous situation that most office workers could relate to would be having to use three different word processors in an average day. Image if you had to be trained on and familiar with all the subtleties of Microsoft Word, Corel WordPerfect and Lotus WordPro!

These problems created by different surveillance systems are being addressed through the definition of **standards for system development activities**. As previously mentioned, creation of **data architecture standards** will ensure, just as it did for the data collection forms, that information is entered and stored in a consistent and uniform way. Having data stored in a uniform way also means that they can more easily be transferred from one system to another so that duplicate data entry is reduced.

Another relevant set of standards has to do with the **user interface** of CDC-developed surveillance systems. A person trained on any one system, for example, should be able to move to another without changing the way they interact with the software. A set of standards for a **common user interface** will guarantee that all systems look and work similarly. It is expected that the same set of **user interface standards** will be applied both to Windows-based applications and to Internet-browser based data collection systems. This type of integration through interface of the web and the stand-alone PC is the same approach that Microsoft is taking with its operating systems and application interfaces.

Data Transmission from Local to State Health Department and to CDC

Once surveillance data are entered into computerized data management systems, they are not only analyzed within the organization to which they were reported, but are also transmitted for analysis at other levels. Simply speaking, **electronic reporting** may occur as data are sent from the health care setting to local (city or county) health departments, then on to State health departments, and finally to CDC. With the current myriad of systems in place, there are many different methods for reporting data. For example, diskettes may be mailed, dial-up modems may be used to connect over public telephone lines, leased telephone lines may provide wide-area network used for reporting, or the public Internet may be used. Currently, just for reporting to CDC, all of these methods are in place. There are also different levels of security in terms of electronic encryption methods that are applied. For example, in a recent inventory, over 73 different surveillance systems developed at CDC sent or received surveillance data electronically. Only 19, however, reported encrypting the data for transmission. While virtually all programs do not consider the encryption of their data an issue because individual person or site identifiers are removed before reporting to the next level, there is at least a small risk that a person could be indirectly identified based on data in these individual records.

There are two coordinated efforts that are addressing this problem. The first is the creation of the "**Health Alert Network** (HAN)" that will use the Internet as a backbone for communicating surveillance data (as well as a host of other information such as surveillance reports, training materials, policy documents, etc.) between health departments and CDC. This system is

expected one day to connect the myriad of local health departments with State and territorial health departments and federal agencies, including CDC, nationwide.

The second part is the **Secure Data Network** (**SDN** – sometimes called the "**secure Internet pipeline**"). This pipeline will provide CDC program areas with a secure method for encrypting and transferring files from a health department to a CDC program application across the Internet. (As an Internet-based system, the **SDN** may be considered to be part of **HAN**, not independent of it.) It will also allow CDC to eliminate the multiple methods of receiving data. In addition, using digital certificates and the MD5 message digest, the **SDN** will provide a consistent, transparent method for authenticating the source and ensuring the integrity of those data. This network will raise the standard on security for most of the surveillance activities now supported by CDC.

Eventually the combination of the **HAN** and the standards that are used for the **Secure Data Network** can be extended to support standardized security beyond just reporting to CDC. They will allow any two or more partners in public health, whether they are heath care providers, clinical labs or local and State health departments, to exchange information without risk of eavesdropping by unauthorized parties.

Data Analysis

Individual program areas at CDC and State and local health departments have, over the years, developed many innovative methods for the analysis of data. For example, recent efforts have led to development of techniques that accurately **detect some changing or unusual patterns** of trends or outbreaks of diseases. In addition, **statistical methods** have been developed to account for the delays in reporting of data from providers to health departments to CDC and, where appropriate, to estimate the true incidence of a disease or condition even when not all cases have been reported.

The tools for implementing these methods have been provided to local and State health departments as part of individual surveillance systems, but in general they are not widely available. The closest thing that CDC has to the universal distribution of analysis tools are those contained in the DOS-based EpiInfo software package, however the DOS version of the application does not include some of the more sophisticated techniques described above.

The issue of how to provide **standardized data analysis tools** will be addressed by NEDSS through the identification, adoption and promotion of **statistical component standards**. Software written to these standards will be able to be used and incorporated into a variety of surveillance systems – not only those developed by the CDC but also those that are being used by local and State health departments. As an example of the application of components, state-of-the-art analytic software would be able to be dropped into other software applications in the same way that spreadsheets, presentation graphics and e-mail components are now a standard part of many systems.

Data Analysis: Linkage

Another common problem is the need to link data collected in different surveillance and information systems. For example, persons responsible for notifiable diseases are interested both in the cases reported by providers and also, whether those cases might be linked to those reported in a laboratory-based system, where there is available species or serotype information that indicates that these cases may be part of an outbreak. Or, persons investigating an increase in the number of cases of a vaccine-preventable disease would be interested in determining whether persons with these illness received a certain type or lot of vaccine, information increasingly available, but in a separate location such as an immunization registry. And persons responsible for maternal and child health programs at the State level have noted that how they define and count cases of infectious disease among children does not match the *Case Definitions for Infectious Conditions under Public Health Surveillance* developed by CDC and the Council of State and Territorial Epidemiologists (CSTE) for notifiable disease surveillance.

This issue is also being addressed by the surveillance **data standards**. The data standards will promote the linkage of data, as appropriate to public health needs, either at the individual patient or record level, or more broadly by place and time. **Having standardized definitions for data elements** will help ensure the correct interpretation of data elements. Having data stored using the same sets of codes will mean that epidemiologists and others needing merged data sets will not need to spend as much time understanding the peculiarities of any one system. Finally, these standards, by ensuring **consistent definitions of data and coding of variables**, will also facilitate the development of State data warehouses and the virtual State data centers envisioned by the National Association for Public Health Statistics and Information Systems (NAPHSIS).

Data Access and Dissemination
The ability to access and disseminate appropriate public health data and information in a timely fashion to targeted audiences is key to making an impact on the population's health. Often, however, there are significant delays in providing access. For example, program areas at CDC often complain that they spend much of their time generating data sets and responses to requests for information by State health departments, other Federal and State agencies, non-profit organizations, the news media, the public, etc. In addition, providing this information typically requires that staff redirect their activities away from other responsibilities. States also point to the same level of resources required to respond to the myriad of organizations within their own area that frequently request public health data and information.

While providing easy access to appropriate public health data and information has been difficult in the past, program areas have also struggled with how to disseminate and/or present data and information results to interested parties. For example, one program area may use the 1990 population census as the denominator for generating rates, while another program area in the same State uses the projected 1998 rates. In addition, program areas may present rates by age in five-year intervals for one disease while disseminating results for another disease using different age ranges, leading to an inability to compare the data (when indicated). Finally, no central location at CDC or in many State health departments exists where people can go to locate these data. This lack of a standardized approach to disseminating data and information at CDC, and

often in State and local health departments as well, impedes the ability of public health professionals to have a direct impact on public health policy and decision-making.

To address these challenges, NEDSS will include the **development of best-practice specifications** for a method to analyze and disseminate data and information, primarily using data warehouses. Through collaboration with people within CDC and State health departments currently developing data warehouses, a method will be developed to solicit specification requirements from potential users, to review available COTS (commercial off-the-shelf) products based on these functionalities, and to provide logical justification for choices, with empirical justification when available and appropriate. Through this process, **a form of a standard off-the-shelf or internally developed software application** will be identified to provide data access capacity to a variety of users with various needs.

Conclusion

As this document illustrates, CDC staff working on NEDSS are focusing on the development, testing, and implementation of standards. These standards will serve as the framework that will support more complete and comprehensive integration of systems in the future. The standards focus on five important areas: **data architecture (data model and data standards)**, **user interface**, **information systems software architecture**, **tools for interpretation, analysis, and dissemination of data**, and **secure data transfer**. While the various systems developed by CDC and State and local health departments will remain distinct from one another, the use of standards will assure that surveillance data may be easily shared, that users familiar with one system can easily use another, and that software can be easily shared across programs. In addition this approach will ensure that State-of-the-art statistical methods are readily available to epidemiologists, and that a single secure method is in place for reporting data to CDC.

These standards are just the first step to achieving the desired level of integration among CDC-developed, as well as State- and locally-developed, surveillance systems. However, they will provide an important degree of integration for the collection, management, transmission, analysis, and dissemination of data that does not currently exist. It is expected that this integration will better support public health professionals in their efforts to improve the health of the populations they serve.